

# A structural mean model to allow for noncompliance in a randomized trial comparing 2 active treatments

KRISTA FISCHER\*

*MRC Biostatistics Unit, Institute of Public Health, University Forvie Site,  
Robinson Way, Cambridge CB2 0SR, UK  
krista.fischer@ut.ee*

ELS GOETGHEBEUR

*Department of Applied Mathematics and Computer Science, Ghent University,  
Krijgslaan 281 S9, B-9000 Ghent, Belgium*

BERNARD VRIJENS

*AARDEX Group, Pharmionic Systems SA, Rue des Cyclistes Frontière 24,  
BE-4600 Visé, Belgium*

IAN R. WHITE

*MRC Biostatistics Unit, Institute of Public Health, University Forvie Site,  
Robinson Way, Cambridge CB2 0SR, UK*

## SUMMARY

We propose a structural mean modeling approach to obtain compliance-adjusted estimates for treatment effects in a randomized-controlled trial comparing 2 active treatments. The model relates an individual's observed outcome to his or her counterfactual untreated outcome through the observed receipt of active treatments. Our proposed estimation procedure exploits baseline covariates that predict compliance levels on each arm. We give a closed-form estimator which allows for differential and unexplained selectivity (i.e. noncausal compliance-outcome association due to unobserved confounding) as well as a nonparametric error distribution. In a simple linear model for a 2-arm trial, we show that the distinct causal parameters are identified unless covariate-specific expected compliance levels are proportional on both treatment arms. In the latter case, only a linear contrast between the 2 treatment effects is estimable and may well be of key interest. We demonstrate the method in a clinical trial comparing 2 antidepressants.

*Keywords:* Causal inference; Randomized-controlled trials; Structural mean models.

## 1. INTRODUCTION

Regulators and stakeholders agree that today's drug evaluation process has become expensive to the point of threatening new drug development (DiMasi *and others*, 2003). At the same time, technological

\*To whom correspondence should be addressed.

developments have raised the hope for more targeted, individualized medicine. The challenge is to learn as much as possible from pivotal randomized studies, for instance, on departures from randomized treatment, including prescribed changes of treatment, noncompliance, and dose–response relationships (Efron and Feldman, 1991) or more accurate estimation of treatment efficacy (the causal effect of the prescribed dose as intended) (Goetghebeur and Lapp, 1997).

Analysis of data on actual treatments can be complex. In phase III trials, a standard intention to treat (ITT) analysis is a good starting point, especially for drug regulators motivated by the desire for a conservative conclusion. Conventional alternatives such as per-protocol analysis risk serious bias since nonadherence to the study protocol (treatment noncompliance) may be related to treatment and outcome. In this paper we focus on compliance-adjusted linear regression analysis, which uses randomization as an instrument (White, 2005; Sommer and Zeger, 1991; Goetghebeur and Lapp, 1997; Vansteelandt and Goetghebeur, 2003). Technically, our approach is an application of the theoretical developments in Robins and others (1994) for parameter estimation in structural mean models (SMMs).

Applications of SMMs have hitherto focused on adjusting for observed experimental exposure in trials with a placebo group or an untreated control group. However, active-controlled trials (ACTs) are increasingly warranted as drug treatment becomes standard for ever more diseases (Ellenberg and Temple, 2000). ACTs present new challenges in comparing efficacy. First, unlike in placebo-controlled trials, in ACTs treatment noncompliance need not attenuate the ITT difference. For example, it is possible to observe an ITT difference when 2 treatments are equally efficacious but one is more commonly discontinued. Second, some ACTs are aimed at demonstrating noninferiority of the new treatment to the standard treatment, not superiority. Noncompliance then exaggerates any evidence of equivalence in outcomes, potentially leading to anticonservative behavior (Jones and others, 1996; Vrijens and Urquhart, 2005) as pointed out in the ICH E10 guideline (International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use, 1999). Third, compliance-adjusted analysis is complicated in ACTs by the lack of a reference group of unexposed participants. Without a directly observed treatment-free reference, even the choice and interpretation of the causal estimand requires attention (Cuzick and others, 1997).

Some authors have tackled compliance adjustment in ACTs by making strong assumptions. Robins (1998) did not exploit the randomization but relied on the untestable no-unmeasured-confounders assumption. Roy and others (2008) added relatively strong untestable parametric assumptions.

This paper studies the potential and limitations of the instrumental variable–based SMM estimator of treatment effects in ACTs. Our case study is a clinical trial of fluoxetine and paroxetine, 2 common antidepressant drugs (Demyttenaere and others, 2008). Like most mental health trials, it is compromised both by noncompliance and by dropout, with nonadherent patients particularly prone to dropout (Dunn and others, 2003).

In Section 2, we formalize the model and its estimating equations for our problem. A closed-form estimator is derived in Section 3. Identifiability conditions are discussed in Section 4. In Section 5, we demonstrate the usage of the estimator on the antidepressant trial. We end with a discussion in Section 6.

## 2. AN SMM FOR THE COMPARISON OF 2 ACTIVE TREATMENTS

Consider a randomized 2-arm trial, where  $n$  patients are randomized to one of 2 active treatments, A or B. We define the following (observed or potential) variables for each patient  $i = 1, \dots, n$ :  $Y_i^A$ ,  $Y_i^B$ : (potential) outcome under assignment to treatment A or B, respectively;  $Y_i^0$ : potential treatment-free outcome;  $C_i^A$ ,  $C_i^B$ : (potential) vectors of treatment compliance summaries observed under assignment to A or B, respectively;  $\mathbf{X}_i$ : a vector of baseline covariates;  $R_i^A$ ,  $R_i^B$ : randomization indicators, with value 1, when patient  $i$  is randomized to A (B) and 0 otherwise. We will also use the notation  $R_i = R_i^A$  and  $Y_i = Y_i^A R_i^A + Y_i^B R_i^B$  for the observed outcome.

An SMM defines the average effects of assignment to  $A$  and  $B$  as the expected differences  $E(Y_i^A - Y_i^0 | \mathbf{X}_i, \mathbf{C}_i^A)$  and  $E(Y_i^B - Y_i^0 | \mathbf{X}_i, \mathbf{C}_i^B)$  and assumes these are known functions  $\gamma^A$  and  $\gamma^B$  of compliance, baseline characteristics, and unknown parameter vectors  $\boldsymbol{\psi}^A$  or  $\boldsymbol{\psi}^B$ , respectively (Goetghebeur and Lapp, 1997). Subtracting postulated effects from corresponding outcomes yields on average the expected treatment-free response, conditional on compliance and baseline characteristics:

$$\begin{aligned} E[Y_i^A - \gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) | \mathbf{C}_i^A, \mathbf{X}_i] &= E[Y_i^0 | \mathbf{C}_i^A, \mathbf{X}_i] \text{ and} \\ E[Y_i^B - \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B) | \mathbf{C}_i^B, \mathbf{X}_i] &= E[Y_i^0 | \mathbf{C}_i^B, \mathbf{X}_i]. \end{aligned} \quad (2.1)$$

It is common to assume the following “exclusion restrictions” hold:

$$\gamma^A(0, \mathbf{X}_i; \boldsymbol{\psi}^A) = 0 \text{ and } \gamma^B(0, \mathbf{X}_i; \boldsymbol{\psi}^B) = 0. \quad (2.2)$$

Thus, if no active treatment is received, expected outcome equals expected treatment-free outcome.

Equations in (2.1) reflect 2 SMMs, equivalent to the models introduced by Goetghebeur and Lapp (1997). The functions  $\gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A)$  and  $\gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B)$  represent average effects of  $A$  or  $B$  assignment in a subgroup of patients at given levels of  $\mathbf{X}$  and  $\mathbf{C}^A$  or  $\mathbf{C}^B$ . The models allow for treatment effect heterogeneity, make no assumptions on the effects of one treatment dose in patients observed to take a different dose and make no assumptions on the association between  $\mathbf{C}^A$  and  $\mathbf{C}^B$ . However, as the  $Y_i^0$ -distribution is not observed, the estimator proposed in Goetghebeur and Lapp (1997) is not directly applicable here.

The main interest lies often in a difference between the 2 treatment effects. Conditioning jointly on  $\mathbf{C}_i^A$  and  $\mathbf{C}_i^B$  leads to the single equation:

$$E[Y_i^A - \gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) | \mathbf{C}_i^A, \mathbf{C}_i^B, \mathbf{X}_i] = E[Y_i^B - \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B) | \mathbf{C}_i^A, \mathbf{C}_i^B, \mathbf{X}_i]. \quad (2.3)$$

Now the difference  $\gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) - \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B)$  has a direct interpretation as the expected outcome difference under  $A$  and  $B$  assignments in a subgroup of patients with given values of  $\mathbf{X}$ ,  $\mathbf{C}^A$ , and  $\mathbf{C}^B$ . As the 2 potential compliance measures are never jointly observed in a parallel trial, such subgroups are not identified.

By taking expectations conditional on  $\mathbf{X}_i$  only, either (2.1) or (2.3) implies

$$E[Y_i^A - \gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) | \mathbf{X}_i] = E[Y_i^B - \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B) | \mathbf{X}_i]. \quad (2.4)$$

As the expectations of the variables appearing on the left are estimable from arm  $A$  and those on the right from arm  $B$  data, (2.4) suggests the following estimation procedure. With  $\boldsymbol{\psi} = (\boldsymbol{\psi}^A; \boldsymbol{\psi}^B)$ , an unbiased estimating equation for  $\boldsymbol{\psi}$  is

$$\sum_{i=1}^n \mathbf{U}_i = \sum_{i=1}^n \mathbf{g}(R_i, \mathbf{X}_i) [H_i(\boldsymbol{\psi}) - q(\mathbf{X}_i)] = \mathbf{0}, \quad (2.5)$$

where  $H_i(\boldsymbol{\psi}) = Y_i - R_i^A \gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) - R_i^B \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B)$ ;  $\mathbf{g}(R_i, \mathbf{X}_i)$  is any vector with  $\dim(\boldsymbol{\psi})$  elements, satisfying  $E[\mathbf{g}(R_i, \mathbf{X}_i) | \mathbf{X}_i] = \mathbf{0}$ , and  $q(\mathbf{X}_i)$  is any function of  $\mathbf{X}_i$ . Efficient estimates are obtained with optimal choices of  $\mathbf{g}$  and  $q$ . Following Robins (1994), the optimal function  $q_{\text{opt}}$  is  $q_{\text{opt}}(\mathbf{X}_i) = E[H_i(\boldsymbol{\psi}_0) | \mathbf{X}_i]$ , with  $\boldsymbol{\psi}_0$  the true (unknown) value of  $\boldsymbol{\psi}$ . A different choice of  $q(\mathbf{X}_i)$  will not introduce bias in the estimating equation, but the variance of the solution  $\hat{\boldsymbol{\psi}}$  is a function of  $\text{Var}[H_i(\boldsymbol{\psi}_0) - q(\mathbf{X}_i) | \mathbf{X}_i]$  and thus depends on how well  $q(\mathbf{X}_i)$  approximates  $H_i(\boldsymbol{\psi}_0)$ . So a better precision of  $\hat{\boldsymbol{\psi}}$  can be achieved by including in  $\mathbf{X}_i$  covariates that are predictive for the potential treatment-free outcome.

Since  $E[H_i(\boldsymbol{\psi}_0) | \mathbf{X}_i] = E[Y_i^0 | \mathbf{X}_i]$  is the expected treatment-free response, it is not directly estimable from the data on either active treatment arm, unlike in the placebo-controlled trial with no contamination

in the control arm (Goetghebeur and Lapp, 1997). However, with a parametric model  $q(\mathbf{X}_i) = q(\mathbf{X}_i; \boldsymbol{\alpha})$ , consistent estimates for  $\boldsymbol{\psi}$  and  $\boldsymbol{\alpha}$  are obtained by simultaneously solving the unbiased estimating equations (2.5) and

$$\sum_{i=1}^n \mathbf{d}(\mathbf{X}_i)[H_i(\boldsymbol{\psi}) - q(\mathbf{X}_i, \boldsymbol{\alpha})] = 0, \quad (2.6)$$

with  $\mathbf{d}(\mathbf{X}_i)$  being any function of  $\mathbf{X}_i$ . By analogy with generalized linear models, we choose  $\mathbf{d}(\mathbf{X}_i) = \frac{\partial q(\mathbf{X}_i, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$ .

Given  $q$ , the optimal function  $\mathbf{g}_{\text{opt}}$  is described in Robins (1994) or Goetghebeur and Lapp (1997):

$$\mathbf{g}_{\text{opt},i}^{w'} = w_{\text{opt},i} \mathbf{g}'_{\text{opt},i} = w_{\text{opt},i} \{E[\partial H_i(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}' | R_i, \mathbf{X}_i] - E[\partial H_i(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}' | \mathbf{X}_i]\}, \quad (2.7)$$

where  $w_{\text{opt},i} = \text{Var}\{H_i(\boldsymbol{\psi}) | \mathbf{X}_i\}^{-1}$  can be dropped if the variance does not change with  $\mathbf{X}_i$ . When  $q$  and  $\mathbf{g}$  have their optimal forms,  $\text{Var}\{n^{1/2}[\hat{\boldsymbol{\psi}}(\mathbf{q}_{\text{opt}}, \mathbf{g}_{\text{opt}})]\}$  attains the ‘‘semiparametric efficiency bound.’’ If the conditional variance of  $H_i(\boldsymbol{\psi})$  given  $\mathbf{X}_i$  were to vary on either treatment arm beyond what is expected for the standard generalized linear model, one would weight (2.6) by the inverse variance.

With  $\tau = E[\partial U_i(\boldsymbol{\psi}_0, \mathbf{g}, q)/\partial \boldsymbol{\psi}']$  and  $\Omega = \text{Var}[U_i(\boldsymbol{\psi}_0, \mathbf{g}, q)]$ , a consistent variance estimator for the estimated parameter vector is derived from the corresponding estimates as  $n^{-1} \hat{\tau}^{-1} \hat{\Omega} \hat{\tau}'^{-1}$ .

### 3. CLOSED-FORM ESTIMATOR FOR A LINEAR SMM

A linear SMM is defined by

$$\gamma^A(\mathbf{C}_i^A, \mathbf{X}_i; \boldsymbol{\psi}^A) = \boldsymbol{\psi}^{A'} \mathbf{Z}_i^A \text{ and } \gamma^B(\mathbf{C}_i^B, \mathbf{X}_i; \boldsymbol{\psi}^B) = \boldsymbol{\psi}^{B'} \mathbf{Z}_i^B, \quad (3.1)$$

where the vectors  $\mathbf{Z}_i^A$  and  $\mathbf{Z}_i^B$  may contain compliance summaries (components of  $\mathbf{C}_i^A$  and  $\mathbf{C}_i^B$ ) and their interactions with baseline covariates. This paper focuses on estimation of  $\boldsymbol{\psi}^A$  and  $\boldsymbol{\psi}^B$  and their contrast  $\boldsymbol{\psi}^A - \boldsymbol{\psi}^B$ , the latter being often of greatest interest, given the trial’s objective of direct comparison between treatments A and B. In the absence of an untreated control group, we will show that inference on the contrast  $\boldsymbol{\psi}^A - \boldsymbol{\psi}^B$  is typically more robust.

Assuming a linear model  $q(\mathbf{X}_i, \boldsymbol{\alpha}) = \boldsymbol{\alpha}' \mathbf{X}_i$  for the dependence of potential treatment-free response on  $\mathbf{X}_i$ , the optimal  $\mathbf{g}$  becomes

$$\mathbf{g}_{\text{opt},i} = \{E(\mathbf{Z}_i^A | \mathbf{X}_i)' [R_i^A - E(R_i^A | \mathbf{X}_i)]; E(\mathbf{Z}_i^B | \mathbf{X}_i)' [R_i^B - E(R_i^B | \mathbf{X}_i)]\}', \quad (3.2)$$

where the expectations can be estimated by regression methods.

Let  $\mathbf{G}, \mathbf{Z}, \mathbf{X}$  be matrices with  $i$ th row  $\hat{\mathbf{g}}_{\text{opt},i}$  (or  $\hat{\mathbf{g}}_{\text{opt},i}^w$ , if the use of weights is indicated),  $(R_i^A \mathbf{Z}_i^{A'})$ ,  $(R_i^B \mathbf{Z}_i^{B'})$  and  $\mathbf{X}_i$ , respectively, and let  $\mathbf{Y}$  be the column vector with  $i$ th element  $Y_i$ . The estimating equations (2.5) and (2.6) then have matrix representation

$$\begin{aligned} \mathbf{G}'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\psi} - \mathbf{X}\boldsymbol{\alpha}) &= 0, \\ \mathbf{X}'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\psi} - \mathbf{X}\boldsymbol{\alpha}) &= 0 \end{aligned} \quad (3.3)$$

and the closed-form solution for  $\boldsymbol{\psi}$  is

$$\hat{\boldsymbol{\psi}} = \{\mathbf{G}' P_X \mathbf{Z}\}^{-1} \mathbf{G}' P_X \mathbf{Y}, \text{ with } P_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}', \quad (3.4)$$

provided all inverse matrices exist. Here,  $\mathbf{I}$  is an  $n \times n$  identity matrix. The parameter vector  $\boldsymbol{\alpha}$  can subsequently be estimated as  $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\psi}})$ . When in  $\mathbf{g}_{\text{opt}}$ ,  $E(\mathbf{Z}_i^A | \mathbf{X}_i)$  and  $E(\mathbf{Z}_i^B | \mathbf{X}_i)$  are estimated

from classical linear regression models, this estimator is equivalent to the 2-stage least squares estimator often used in econometrics and social science applications (Maracy and Dunn, 2010). A consistent estimator for the variance–covariance matrix of the estimated parameter vector  $\boldsymbol{\psi}$  can be derived from (3.4), as

$$\widehat{\text{Var}}(\hat{\boldsymbol{\psi}}) = \{\mathbf{G}' P_X \mathbf{Z}\}^{-1} \mathbf{G}' P_X \mathbf{G} \{\mathbf{G}' P_X \mathbf{Z}\}^{-1} \hat{\sigma}^2. \quad (3.5)$$

Here,  $\hat{\sigma}^2$  is the estimated variance of  $H_i(\boldsymbol{\psi})$  (given  $\mathbf{X}_i$ ), obtained as the residual variance:  $\hat{\sigma}^2 = \sum_{i=1}^n e_i^2 / df$  with  $df = n - \dim \boldsymbol{\psi} - \dim \boldsymbol{\alpha}$  and residuals  $e_i$  defined by  $e_i = H_i(\hat{\boldsymbol{\psi}}) - \mathbf{q}(\mathbf{X}_i, \hat{\boldsymbol{\alpha}}) = Y_i - R_i^A \mathbf{Z}_i^A \hat{\boldsymbol{\psi}}^A - R_i^B \mathbf{Z}_i^B \hat{\boldsymbol{\psi}}^B - \hat{\boldsymbol{\alpha}}' \mathbf{X}_i$ . The estimator is easy to implement in software supporting simple matrix language. The main practical issue is identifiability of the parameters, which corresponds mathematically to the existence of inverse matrices in (3.4). We discuss this next.

## 4. IDENTIFIABILITY

### 4.1 Condition for identifiability of the 2 distinct parameters in a simple linear SMM

Consider first univariate compliance summaries: (potential) exposures  $C_i^A$  and  $C_i^B$ , for example, the percentage of the assigned dose of drug that would be taken by patient  $i$  if assigned to treatment A or B, respectively. We assume  $C_i \geq 0$ , although the methodology can accommodate negative values. The values  $C_i = 0$  have here the special meaning of absence of treatment exposure. For a simple linear SMM, estimating equations are derived from

$$E[Y_i^A - \psi^A C_i^A | \mathbf{X}_i] = E[Y_i^B - \psi^B C_i^B | \mathbf{X}_i]. \quad (4.1)$$

Note that this SMM implies for the ITT treatment effect:  $E(Y_i^A - Y_i^B) = \psi^A E(C_i^A) - \psi^B E(C_i^B)$ . In the special case where  $E(C_i^A) = E(C_i^B) = \mu_c$ , for instance, when compliance can be seen as an attribute of the patient (as in Efron and Feldman, 1991), the ITT treatment difference is proportional to the difference of the 2 SMM parameters:  $E(Y_i^A - Y_i^B) = (\psi^A - \psi^B) \mu_c$  but this is not generally true. In general, the parameters  $\psi^A$  and  $\psi^B$  are identified under the following condition:

“The two distinct parameters of a simple linear SMM (4.1) comparing each active treatment with treatment free response are identified, unless there exists a constant  $k$  such that  $E(C_i^B | \mathbf{X}_i) = k E(C_i^A | \mathbf{X}_i)$ . In the latter case, only the contrast  $\delta = \psi^A - k \psi^B$  is estimable.”

One hence needs baseline covariates which predict compliance differently across randomized groups. This would hold if the 2 treatments were preferred by different patient subcategories.

A detailed discussion on estimation when the 2 separate parameters cannot be estimated, with some practical recommendations, can be found in Section A of the Supplementary Material available at *Biostatistics* online.

### 4.2 Identifiability of a multiple linear SMM

For some purposes, a simple linear SMM is an oversimplification and model (3.1) would include the causal effect of multiple exposures or several summaries of dose timing (possibly interacting with certain baseline covariates). For all distinct parameters to be identified, it is then necessary and sufficient that the matrix  $E(\mathbf{Z} | \mathbf{X})$  with  $i$ th row  $E(\mathbf{Z}_i | \mathbf{X}_i) = E[(\mathbf{Z}_i^A, \mathbf{Z}_i^B)' | \mathbf{X}_i]$  is of full column rank. This can happen as long as  $\dim(\mathbf{X}_i) \geq \dim(\mathbf{Z}_i) - 1$  and none of the components of  $E(\mathbf{Z}_i | \mathbf{X}_i)$  can be expressed as a linear combination of the others. Even then, linear or near-linear dependencies in  $E(\mathbf{Z} | \mathbf{X})$  can make causal parameter estimates highly variable. They might nevertheless allow for precise estimation of useful contrasts such as  $\delta$ .

### 4.3 Simulations

We conducted simulations to study the precision and possible finite sample bias of the estimates, assuming the model (4.1), also examining the performance when parameters are unidentified. Details of simulations are given in Section B of the Supplementary Material available at *Biostatistics* online.

The results indicate that estimator behaves quite well when the 2 expected compliance measures, given baseline characteristics, are different with correlation 0.75 or even 0.95, with a relatively small finite sample bias in small samples. When the correlation between the 2 expected compliance summaries is 1, but the 2 summaries differ by an additive constant, the estimate is still reasonably precise for sample size of 2,000 and might also work for  $n = 400$ . When the 2 summaries differ by a multiplicative constant or are equal, the estimates are more imprecise and in these cases one is unlikely to identify the 2 distinct parameters.

The simulations demonstrate that the difference  $\delta = \psi^A - \psi^B$  is always estimated with much better precision than the 2 distinct parameters  $\psi^A$  and  $\psi^B$ , the SMM algorithm providing unbiased and relatively precise estimates even when the 2 expected compliance summaries are equal.

### 4.4 Extensions

The linear SMM defined by (3.1) can be applied in a much wider range of applications than the simple situation, where on each arm only one treatment, A or B, is available and the exclusion restriction (2.2) holds. Section C of the Supplementary Material available at *Biostatistics* online discusses 2 possible extensions. First, one can allow for contamination—treatment A being received by some patients on arm B and treatment B by some patients on arm A, which leads to a special case of an SMM with multivariate compliance summaries, possibly still simplified to a 2-parameter model. Second, one can relax the exclusion restriction by allowing a constant nonzero difference between potential outcomes on arms A and B in addition to the coefficients of the received dose. In this case, the SMM involves 3 causal parameters that may be difficult to identify in practice with reasonable precision.

## 5. A TRIAL OF 2 ANTIDEPRESSANTS

We apply our method to compare 2 antidepressant treatments, fluoxetine and paroxetine, accounting for variability in drug exposure due to suboptimal compliance with prescribed therapy.

The trial (Demyttenaere and others, 2004, 2008) was designed as part of a study to develop an antidepressant compliance questionnaire. It was a double-blind randomized multicentre study in patients with major depressive disorder. 85 patients were randomized to fluoxetine ( $n = 42$ , later referred to as treatment A) or paroxetine ( $n = 43$ , treatment B), 20 mg/day, for 22 weeks after an initial wash out and run-in period on placebo. The severity of depression was assessed by the Hamilton depression scale at clinic visits before the run-in period (Visit 1), at randomization (Visit 2), after the initial 2 weeks on randomized treatment (Visit 3), and then every 4 weeks for the next 5 months (Visits 4–8). Medication event monitoring systems (MEMS) were used to automatically compile drug dosing histories (Vrijens and others, 2005); treatment compliance was the percentage of prescribed dose actually taken by the patient during a given time period.

As all patients returned their MEMS devices, complete compliance data were available. Also, Hamilton scores at Visits 1 and 2 were available for all 77 patients, but 3, 2, 4, 3, 4, and 8 patients dropped out prior to Visits 3–8, respectively, and so the Hamilton score at final Visit 8 was available for only 53 patients (69%).

Figure 1 shows the evolution of individual Hamilton scores during the trial. The ITT analysis reveals no significant difference between the 2 treatment arms, although both arms saw significant reduction in average Hamilton scores over the active period (45% at 6 weeks and 76% at 22 weeks). Estimated

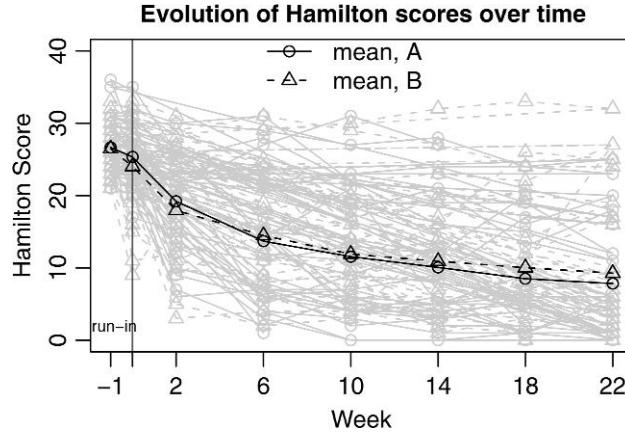


Fig. 1. Evolution of Hamilton scores during the 23 weeks of the trial.

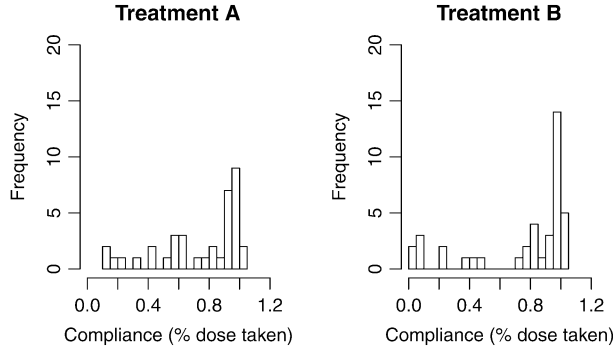


Fig. 2. Histograms of compliance, estimated as percentage of prescribed dose taken by the time of Visit 8.

treatment compliance varied widely between patients (Figure 2). Here, we use the SMM methodology to explore how drug effects compare at different drug-specific compliance levels. To handle dropouts in the data set, we used multiple imputation (Rubin, 1987) assuming the data are missing at random.

### 5.1 Possible SMMs for the antidepressant trial

Let  $C_i(j, k)$  denote the percentage of prescribed pills taken by patient  $i$  between Visits  $j$  and  $k$  and  $C_i^w(j)$  the percentage of prescribed pills taken during the week before Visit  $j$ . Suppose  $\mathbf{C}_i^A$  and  $\mathbf{C}_i^B$  are vectors with components  $C_i(j, k)$  and  $C_i^w(j)$  for all  $j$  and  $k$ . We assume the following models for the final (Visit 8) outcomes  $Y_i^A(8)$  and  $Y_i^B(8)$ :

$$\begin{aligned} E[Y_i^A(8) - Y_i^0(8) | \mathbf{C}_i^A, \mathbf{X}_i] &= \psi_1^A C_i^A(2, 8) + \psi_2^A C_i^{w,A}(8), \\ E[Y_i^B(8) - Y_i^0(8) | \mathbf{C}_i^B, \mathbf{X}_i] &= \psi_1^B C_i^B(2, 8) + \psi_2^B C_i^{w,B}(8). \end{aligned} \quad (5.1)$$

These models assume that the effect of each treatment comprises a cumulative effect of the total dose taken since baseline, reflected by  $\psi_1^A$  ( $\psi_1^B$ ), and a short-term effect  $\psi_2^A$  ( $\psi_2^B$ ). We fit this model and 4 nested models, with either  $(\psi_1^A, \psi_1^B)$ ,  $(\psi_2^A, \psi_2^B)$ ,  $(\psi_2^A, \psi_1^B)$ , or  $(\psi_1^A, \psi_2^B)$  being set to equal  $(0, 0)$ , to explore whether one of these effects may be dominant.

## 5.2 A goodness of fit test

A test for goodness of fit (GOF) is based on

$$E(H_i|\mathbf{X}_i, R_i^A = 1) = E(H_i|\mathbf{X}_i, R_i^B = 1) = E(Y_i^0|\mathbf{X}_i).$$

If the parameter estimates  $\boldsymbol{\psi}^A$  and  $\boldsymbol{\psi}^B$  are close to the true values, then the estimates  $\hat{H}_i = Y_i - R_i^A \hat{\boldsymbol{\psi}}^A \mathbf{Z}_i^A - R_i^B \hat{\boldsymbol{\psi}}^B \mathbf{Z}_i^B$  should have similar expectation, given  $\mathbf{X}$ , in both randomized groups. So, as a (partial) GOF test, one can test for the interaction effect of  $R$  and  $\mathbf{X}$  in a linear regression model for  $\hat{H}$ . In complete data, one can use the analysis of variance  $F$ -test comparing regression models for  $\hat{H}$  on  $X$  and  $R$ , with and without the interaction. The numerator and denominator degrees of freedom for the  $F$ -statistic are  $p - q$  and  $n - (1 + 2p - q)$ , where  $p$  is the number of components in  $\mathbf{X}_i$  and  $q$  is the number of estimated SMM parameters. In the presence of missing data, the testing procedure and degrees of freedom were modified using the formulas by Reiter (2007).

## 5.3 Results

First, a set of baseline covariates was selected based on the results of separate multiple regression models for compliance summaries and final Hamilton scores on both arms. Details on the covariate selection and multiple imputation procedure are given in Section D of the Supplementary Material available at *Biostatistics* online.

The resulting estimates for SMM parameters and their standard errors are presented in Table 1. According to the Wald test, all models except Model 5 provide either significant or borderline significant evidence against the null hypothesis that all causal parameters are 0, the  $p$ -value for Model 4 being below the conventional  $\alpha$ -level of 0.05. Model 3 gives the best fit according to the  $F$ -test but the parameter estimates have high variability and the model is hard to interpret. Model 4 is the most convincing model and suggests that treatment A has a more cumulative effect (so  $C^A(2, 8)$  is important), whereas B has a more short-term effect (so  $C^{w,B}(8)$  is important)—both parameter estimates being significantly different from 0. The results are in accordance with the knowledge that fluoxetine (A) has significantly longer half-life than paroxetine (B) (Rosenbaum and others, 1998).

The estimated treatment contrast for full compliers (both compliance summaries being equal to 1) is estimated with better precision than the distinct parameters. In the best-fitting Models 3 and 4, the hypothesis of no treatment difference for full compliers cannot be rejected. Model 4 estimates a significant effect of each separate treatment relative to the treatment-free reference, as defined here. Since Figure 2 reveals how few observations directly reflect the (near) treatment-free response, the latter result is model dependent and needs to be interpreted with caution.

Table 1. Estimated SMM parameters (standard error), treatment contrast (standard error) for full compliers, and GOF tests for 5 SMMs and the null model

	Coefficient				Contrast for full compliers	Test for $\boldsymbol{\psi} = 0$ ( $p$ -value)	GOF $F$ ( $p$ -value)
	$C^A(2, 8)$	$C^{w,A}(8)$	$C^B(2, 8)$	$C^{w,B}(8)$			
Null	—	—	—	—	—	—	1.41 (0.20)
SMM 1	-14.7 (7.1)	—	-9.3 (6.9)	—	-5.4 (2.9)	2.8 (0.07)	0.96 (0.47)
SMM 2	—	-11.5 (4.6)	—	-5.7 (4.2)	-5.7 (3.5)	3.2 (0.05)	0.93 (0.50)
SMM 3	-12.7 (25.2)	5.6 (18.4)	22.2 (22.0)	-31.7 (15.7)	2.3 (5.9)	2.3 (0.07)	0.23 (0.97)
SMM 4	-15.9 (5.4)	—	—	-13.4 (6.0)	-2.5 (3.3)	4.3 (0.02)	0.51 (0.84)
SMM 5	—	-8.6 (5.1)	-1.5 (4.5)	—	-7.7 (3.3)	2.3 (0.11)	1.32 (0.25)



In sensitivity analyses, we omitted one or more covariates from the causal analysis. This caused standard errors of the estimated causal effects to increase, but the results remained consistent with those tabulated.

The validity of the SMM approach in the presence of missing data (assuming the data is missing at random) was also tested in our simulation study (Section B of the Supplementary Material available at *Biostatistics* online). As expected, missing data led to somewhat increased mean squared error of the estimates using multiple imputation, but no biases are observed.

## 6. DISCUSSION

We have developed SMMs, which provide inferences about the causal effects of treatment, for trials where different arms are assigned different active treatments, in the presence of treatment noncompliance.

We have considered flexible causal models involving several causal parameters. Precise estimation of such models requires the existence of strong baseline predictors for observed exposure on each arm, and the data requirements increase with model dimension. It is common in randomized trials to record the covariates that are known to be associated with the outcome, but the present work highlights the important role of covariates that predict treatment compliance. Ideally, the covariates should produce different predictions for the 2 compliance summaries, as for example, when the treatments differ by aspects that are important to the patient, such as half-life, side effects, frequency of dosage, etc.

Our models assume that the baseline covariates do not modify the causal effect of either treatment. Interaction terms between compliance and baseline covariates could be included in the vector  $\mathbf{Z}$  (3.1), as was done in placebo-controlled trials (Goetghebeur and Lapp, 1997). However, this cannot be done for all baseline covariates: the parameters are only identified when enough baseline covariates are (differentially) predictive for but are assumed not to freely interact with compliance on arms A or B.

From a clinical perspective, the ability of baseline covariates to identify patients whose compliance is likely to differ under the 2 assignments can help to guide treatment decisions. Our models can be used to estimate the benefit of assignment to treatment A rather than treatment B for a particular patient's set of baseline covariates.

Interpretation of parameters such as  $\psi^A$  and  $\psi^B$  requires care since different causal parameters defined in this paper pertain to different subpopulations. If one starts from (2.1) then  $\psi^A c^A$  is the average causal effect of compliance  $c^A$  compared with no compliance in the subpopulation with compliance  $c^A$ . Equation (2.3), which conditions on the unobserved joint distribution of  $C_i^A$  and  $C_i^B$ , relies on a stronger assumption which implies that  $\psi^A c^A - \psi^B c^B$  is the ITT difference (the expected difference in outcome between assignment to A and assignment to B) for the subpopulation  $(C^A, C^B) = (c^A, c^B)$ . The practical usefulness of this interpretation is limited as long as we have no information on how patients would comply with both treatment assignments. Thus, for example, if  $C^A$  and  $C^B$  are defined as percentages of assigned dose then  $\psi^A - \psi^B$  is the ITT difference in the "perfect compliers" subgroup with  $C^A = C^B = 1$ . However, such a subgroup may not exist or may be hard to identify if it does. It may be more realistic to consider the ITT difference in a subgroup defined by (for example)  $0.8 \leq (C^A, C^B) \leq 1.2$ .

Equations (2.1) involve a counterfactual untreated outcome  $Y^0$ . In some settings, some patients could not ethically receive no active treatment, so our use of  $Y^0$  is questionable. Models of form (2.3) avoid this problem.

In the depression trial analyzed here, (2.3) requires the causal effect of compliance with fluoxetine to be unaffected by the potential compliance of the same patients with paroxetine (and vice versa). There is no obvious reason to expect such interaction, so we interpret the estimated contrast  $\psi^A - \psi^B$  as the expected ITT difference for those patients whose compliance under either assignment would be close to 1.

The causal parameter estimates in the best-fitting SMM seem to reflect known differences in the duration of action of the 2 treatments.

We close by reminding readers that the causal model framework provides some prior structure which our data alone can neither refute nor confirm. Nevertheless, in many problems, the linear SMM is a good first-order approximation to the true average causal effect.

#### SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

#### ACKNOWLEDGMENT

*Conflict of Interest:* None declared.

#### FUNDING

UK Medical Research Council (U.1052.00.006) to K.F., I.R.W.; Interuniversity Attraction Poles research network grant P6/03 of the Belgian government (Belgian Science Policy) to E.G.

#### REFERENCES

- CUZICK, J., EDWARDS, R. AND SEGNAV, N. (1997). Adjusting for non-compliance and contamination in randomized controlled trials. *Statistics in Medicine* **16**, 1017–1029.
- DEMYTTENAERE, K., ADELIN, A., PATRICK, M., WALTH'ERE, D., KATRIEN DE, B. AND MICH'ELE, S. (2008). Six-month compliance with antidepressant medication in the treatment of major depressive disorder. *International Clinical Psychopharmacology* **23**, 36–42.
- DEMYTTENAERE, K., BRUFFAERTS, R., ALBERT, A., MESTERS, P., DEW, W., DEBRUYCKERE, K. AND SANGELEER, M. (2004). Development of an antidepressant compliance questionnaire. *Acta Psychiatrica Scandinavica* **110**, 201–207.
- DI'MASI, J. A., HANSEN, R. W. AND GRABOWSKI, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22**, 151–185.
- DUNN, G., MARACY, M., DOWRICK, C., AYUSO-MATEOS, J. L., DALGARD, O. S., PAGE, H., LEHTINEN, V., CASEY, P., WILKINSON, C., VAZQUEZ-BARQUERO, J. L. and others (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *The British Journal of Psychiatry* **183**, 323–331.
- EFRON, B. AND FELDMAN, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association* **86**, 9–17.
- ELLENBERG, S. S. AND TEMPLE, R. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. part 2: practical issues and specific cases. *Annals of International Medicine* **133**, 464–470.
- GOETGHEBEUR, E. AND LAPP, K. (1997). The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Journal of the Royal Statistical Society, Series C* **46**, 351–364.
- INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE (1999). Choice of control group and related issues in clinical trials (E10). *Technical Report*. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. <http://www.ich.org/LOB/media/MEDIA486.pdf>.
- JONES, B., JARVIS, P., LEWIS, J. A. AND EBBUTT, A. F. (1996). Trials to assess equivalence: the importance of rigorous methods. *British Medical Journal* **313**, 36–39.

- MARACY, M. AND DUNN, G. (2010). Estimating dose-response effects in psychological treatment trials: the role of instrumental variables. *Statistical Methods in Medical Research* (in press: first published online on November 26, 2008). doi: 10.1177/0962280208097243.
- REITER, J. P. (2007). Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrika* **94**, 502–508.
- ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics—Theory and Methods* **23**, 2379–2412.
- ROBINS, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine* **17**, 269–302.
- ROBINS, J. M., ROTNITZKY, A. AND ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.
- ROSENBAUM, J. F., FAVA, M., HOOG, S. L., ASCROFT, R. C. AND KREBS, W. B. (1998). Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. *Biological Psychiatry* **44**, 77–87.
- ROY, J., HOGAN, J. W. AND MARCUS, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9**, 277–289.
- RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- SOMMER, A. AND ZEGER, S. L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.
- VANSTEELENDT, S. AND GOETGHEBEUR, E. (2003). Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society, Series B* **65**, 817–835.
- VRIJENS, B., TOUSSET, E., RODE, R., BERTZ, R., MAYER, S. AND URQUHART, J. (2005). Successful projection of the time course of drug concentration in plasma during a 1-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *Journal of Clinical Pharmacology* **45**, 461–467.
- VRIJENS, B. AND URQUHART, J. (2005). Patient adherence to prescribed antimicrobial drug dosing regimens. *Journal of Antimicrobial Chemotherapy* **52**, 616–627.
- WHITE, I. R. (2005). Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research* **14**, 327–347.

[Received July 9, 2009; revised July 20, 2010; accepted for publication July 22, 2010]